Ranking Firms Using Revealed Preference

Online Appendix

Isaac Sorkin

A   Appendix: Constructing datasets

Being able to track employers over time is central to measuring employer-to-employer flows and administrative errors in the employer identifiers would lead to an overstatement of flows. Following Benedetto et al. (2007), I assume that large groups of workers moving from employer A to employer B in consecutive periods—especially if employer B did not previously exist—likely reflects errors in the administrative data rather than a genuine set of flows. As such, I correct the employer identifiers using worker flows. I use the Successor-Predecessor File and assume that if 70% or more of employer A's workers moved to employer B, then either 1) employer B is a relabelling of employer A or 2) employer B acquired employer A. Therefore, I do not count such "moves" as employer-to-employer transitions.

A.A   Annual dataset

I follow Abowd et al. (2003) to construct the dataset to estimate the earnings decomposition. I depart from them to define employment in a way that is consistent with how employment is defined to construct employer-to-employer flows, to follow more recent literature in imposing age restrictions, and to follow more recent literature in dropping jobs with very low earnings.

For the purposes of estimating the earnings decomposition, the *annual dominant employer* is the employer from which the worker had the highest earnings in the calendar year. This employer is chosen from the employers from which the worker had received earnings for two or more consecutive quarters within the calendar year; the reason to make this restriction is to allow me to code transitions between employers as employer-to-employer or employer-to-nonemployment-to-employer.[1] In this set of jobs, the annual dominant employer is the one with the highest total earnings in the calendar year.

To construct annualized earnings, for each quarter within a year I first identify the nature of the worker's attachment to the employer. Specifically, code quarter t of earnings into one of the following two mutually exclusive categories: full-quarter (if earnings from the employer are in quarters $t-1$, t and $t+1$) or continuous (if earnings are in quarters $t-1$ and t or in t and $t+1$). Annualize these earnings as follows. First, if the worker had any quarters of full-quarter earnings, take the average of these quarters and multiply by 4 to get an annualized salary. Second, if the worker did not have full-quarter earnings and has any quarters of continuous earnings, take the average of these and multiply by 8 to get an annualized salary. The justification for this procedure is that if a worker is present in only two consecutive quarters and if employment duration is uniformly distributed then on average the earnings represent $\frac{1}{2}$ a quarter's work, while if a worker is present in both adjacent quarters then the earnings reflect a full quarter's work.[2] Then take the log of these earnings.

I then make two additional sample restrictions. First, I keep workers aged 18-61 (on December 31st of the year), inclusive. This is an attempt to avoid issues with retirement. This age restriction is similar to, e.g., Card et al. (2013) (20-60 in Germany) and Taber and Vejlin (2016) (19-55 in Denmark), though Abowd et al. (2003) do not report imposing any age restriction. Second,

---

[1]This eliminates quarters of employment that Abowd et al. (2003, pg. 15-16) term "discontinuous," that is, where a worker is observed in neither adjacent quarter. Abowd et al. (2003, pg. 15-16) report that such discontinuous quarters of employment accounted for 5 percent of person-year observations in their final dataset. Second, it eliminates "continuous" quarters of employment where the first quarter of the match is quarter IV within the year, and the second quarter is quarter I of the following year. Under the assumption that continuous quarters are uniformly distributed within the year, this eliminates $\frac{1}{8}$ of continuous workers. Abowd et al. (2003, pg. 15-16) report that continuous quarters account for 11 percent of observations in their final dataset, so this eliminates about 1.4 percent of observations.

[2]In the small number of cases where a worker had forward-looking continuous employment in quarter IV as well as another quarter of continuous employment at the same employer, I included this quarter in the earnings calculation.

following Card et al. (2013) I drop observations with *annualized* earnings of less than \$3,250 in 2011:IV dollars.[3]

I now summarize how the various sample restrictions affect the same size. Table A1 in Appendix L shows that there are about 650 million person-employer-years before imposing an earnings test, 614 million after imposing an earnings test, and 505 million after going down to one observation per person per year. This means that after dropping the low-earnings jobs, there are an average of 1.2 employers per person per year.[4] Table A2 shows the distribution of the number of jobs per year in row 2 of Table A1.

Table A3 shows that on the full annual dataset, 91% share of person-year observations are full-quarter and 9% are continuous.[5] Table A4 shows the distribution of the number of years per person. About 40% of the people are in the dataset for all 7 years, and only 13% are in the dataset for only a single year. Table A5 shows that there is a substantial amount of mobility in this sample: half of the workers have two or more employers. Table A6 shows that about 10% of person-employer matches (or 30% of person-years) last for the entire span of my data. However, almost half of matches (20% of person-years) only last for a single year.

A.B Quarterly dataset

I build on ideas developed in Bjelland et al. (2011) and Hyatt et al. (2014). Specifically, the procedure of restricting to jobs with two quarters of earnings and using overlapping quarters of earnings to label an employer-to-employer transition comes from Bjelland et al. (2011, pg. 496, equation 2). The idea of using earnings in the two quarters to select the dominant job is found in Hyatt et al. (2014, pg. 3).

For the purposes of measuring flows, the *quarterly dominant employer* in quarter t is the employer from which the worker had the highest earnings summing over quarter t and quarter $t - 1$. This job is chosen from among the employers where the worker had positive earnings in *both* quarter t *and* quarter $t - 1$. To count as employment, the earnings must pass the the same earnings test as for the annual dataset.[6]

For the person-quarters that remain after the earnings test, the goal is to select a single employer—the quarterly dominant employer. The quarterly dominant employer is the employer from which the worker has the most total earnings summing across $t - 1$ and t. There is one exception to this selection rule. If a worker has earnings from her *annual* dominant employer in quarters $t - 1$ and t, then this employer is the quarterly dominant employer regardless of whether it is the employer with the most total earnings summing across $t - 1$ and t. The reason for prioritizing the annual dominant job is that I want to use this quarterly dataset to code transitions between annual dominant jobs so it is important that they appear in the quarterly dataset.

If a worker has different quarterly dominant employers in quarter t and quarter $t + 1$, then this worker had earnings from both employers in quarter t and I label the worker as having undergone an employer-to-employer transition in quarter t. If a worker has no dominant employer in quarter

---

[3]Card et al. (2013) drop daily wages of less than 10 euros. 10 euros $\times \approx 1.3$ euros per dollar $\times 250$ days per year= 3,250.

[4]For Germany, Card et al. (2013, Appendix Table 1a, row 5) find 1.10 employers per person per year, and this number is stable from 1985 through 2009.

[5]Abowd et al. (2003, pg. 15-16) find 84% are full-quarter, 11% are continuous, and 5% are discontinuous.

[6]Sum together the two quarters of earnings and multiply by 4. If the earnings are below \$3,250, then drop the person-employer match. Multiplying by 4 is justified if one assumes that each quarter is a continuous quarter of employment. The assumption that this is a continuous quarter of employment does lead to more jobs being included than the annual dataset; specifically, if a job is actually full-quarter, then the annualized earnings treating it as full quarter can be lower than the annualized earnings assuming it is a continuous quarter.

$t+1$, then, with one exception highlighted a little later in this section, I consider that worker to have been nonemployed in quarter $t = 1$, so I label the transition from the quarter t dominant employer as a transition into nonemployment.[78]

I depart from prior work to address the possibility that workers move on the seam between two quarters (Hyatt and McEntarfer (2012) emphasize that on some outcomes these transitions look like employer-to-employer moves). To make this concrete, suppose that I observe a worker at firm A in quarter $t - 2$ and $t - 1$ and at firm B in t and $t + 1$. Then the definitions developed previously say that in quarter $t - 1$ firm A is the dominant employer and in quarter $t + 1$ firm B is the dominant employer. But in quarter t the worker had no dominant employer because it was not the second consecutive quarter of any employment relationship. So the transition from A to B was an employer-to-nonemployment-to-employer transition. It might be, however, that the worker's last day at A was the last day of quarter $t - 1$ and her first day at B was the first day of quarter t, so this was actually an employer-to-employer transition. The way I attempt to capture these transitions is to use the stability of earnings across quarters to suggest that a worker was probably employed for the full quarter in both quarters. So, if the earnings from firm A in quarters $t - 2$ and $t - 1$ are within 5% of each other (using quarter $t - 1$ earnings as the denominator), then this employer is the dominant employer in quarter t. This then allows me code the transition from A to B as employer-to-employer. Table A7 shows that this correction accounts for 3.5% of the employer-to-employer transitions in my dataset.

The final result is a dataset that at the quarterly level says where the person was employed and, if this is a new job, says whether the worker came to this job directly from another job, or had an intervening spell of nonemployment.

A.C   Combining the quarterly and annual datasets

The goal of combining the datasets is to use the detail of the quarterly dataset to label each transition between annual dominant employers as an employer-to-employer or an employer-to-nonemployment-to-employer transition.

To label the transition as employer-to-employer or employer-to-nonemployment-to-employer, I proceed as follows. First, identify consecutive observations where a worker has a different annual dominant employer; to be concrete, suppose that the worker's annual dominant employer is A in 2002 and B in 2003.[9] Second, look at the quarterly dataset and find the last quarter that the worker is employed at A (this might be in 2002 or 2003). Third, look at the quarterly dataset and find the first quarter that the worker is employed at B (this might be in 2002 or 2003). If the last quarter at A and first quarter at B are adjacent, then there was an overlapping quarter of earnings and I label this an employer-to-employer transition. If not, then typically I label this an employer-to-nonemployment-to-employer transition. The exception to labelling the transition an employer-to-nonemployment-to-employer transition is if the worker made an employer-to-employer move through some third (and possibly fourth or fifth) employer en route to moving from A to B. Suppose, for example, that the worker makes the following transitions (where EE is employer-to-employer and ENE is employer-to-nonemployment-to-employer): $A \xrightarrow{EE} C \xrightarrow{EE} B$. Because the worker only made employer-to-employer transitions between A and B, I label this an employer-

---

[7]Similarly, Burgess et al. (2000) drop matches that only last a single quarter.

[8]This definition of a transition into nonemployment will pick up very few recalls as employer-to-nonemployment-to-employer transitions. The reason is that even if a worker is nonemployed awaiting recall for 13 weeks, the probability that I record a quarter with zero earnings from her employer is less than 10% ($\frac{1}{13}$).

[9]It is possible that a worker only appears in the annual dataset in nonconsecutive years—say, 2002 and 2004. In this case, the procedure ends up labelling the transition an employer-to-nonemployment-to-employer.

to-employer transition between annual dominant employers. Alternatively, suppose that I observe $A \xrightarrow{EE} C \xrightarrow{ENE} B$. Then I label the transition between annual dominant employers an employer-to-nonemployment-to-employer transition.

If a worker never has another employer, then I do not attempt to label this transition. For example, if a worker has a dominant employer in 2006 and no dominant employer in 2007, then I do not record a separation in 2006. The reason is that this could occur for any number of reasons: 1) a worker ages out of my age range, 2) a worker moves out of my states, or 3) a worker leaves the labor force.

B   Appendix: Additional evidence on earnings changes

This appendix provides additional evidence on how the earnings changes line up with the firm effects. As emphasized by Chetty et al. (2014), a measure of bias in firm effects (or, in their case, teacher value-added) is to consider the β_1 coefficient in the following regression:

$$(1) \qquad y^r_{i,t} - y^r_{i,t-1} = \beta_0 + \beta_1 \left[\tilde{\hat{\Psi}}_{\boldsymbol{J}(i,t)} - \tilde{\hat{\Psi}}_{\boldsymbol{J}(i,t-1)} \right] + \epsilon_{i,t}, \forall \text{ i,t s.t. } \boldsymbol{J}(i,t) \neq \boldsymbol{J}(i,t-1),$$

where $y^r_{i,t} = y_{i,t} - x'_{it}\hat{\beta}$ is the residualized earnings, and $\tilde{\hat{\Psi}}_{\boldsymbol{J}(i,t)}$ is the shrunken firm effect. If the firm effects are unbiased, then we expect $\hat{\beta}_1 = 1$. The top panel of Figure A2 shows that this is the case. The figure plots 20 bins of changes in firm effects at all transitions between annual dominant employers against the average individual-level change in earnings on these transitions. The solid line plots the best-fitting line from a regression run on the individual-level data. The thin-dashed line shows the line that would be expected if the firm effects were unbiased. The lines are identical and the coefficient is 1.005. The bottom panel shows the analogous figure for the EE transitions, and the slope is 0.813. Formally this finding could be interpreted as indicating misspecification, though it is not clear whether the departure is quantitatively important. Figure A4 reports the results of a conceptually similar exercise where, following Card et al. (2013, pg. 997), I plot event studies around transitions from lower- to higher-paying firms and vice-versa and show that earnings change in opposite directions with equal magnitudes.

While it may seem mechanical that the firm effects would predict the individual-level changes, this finding does not hold if the AKM decomposition is seriously misspecified. To show this, I simulate data from a model where mobility is on the basis of the comparative advantage (e.g., Eeckhout and Kircher (2011), Lopes de Melo (2016) and Hagedorn et al. (2017)). This type of model implies that the residual plays a large role in determining mobility and generates at least two implications which are at odds with the data. First, there are no earnings cuts on EE transitions, and the individual-level earnings changes always lie above the x-axis. Second, there is not the approximate symmetry in earnings changes from moving to a better or a worse firm (Card et al. (2013, pg. 990) emphasize this symmetry property). Figure A3 plots the analogous figure to Figure A2b with data simulated from the example production function in Eeckhout and Kircher (2011). The estimate of β_1 is about 0.4, and unlike in the data, the earnings changes display a v-shape in the firm effects changes. The v-shape comes from earnings increases accruing to workers whose comparative advantage is working at the lowest productivity firms.

C   Appendix: Selection-correcting the earnings

I selection-correct the earnings equation by combining the proportionality assumption and the results of the search model. That is, I add the expectation of the error term from the search model to the earnings equation.

In the first period of a worker's employment relationship, this expectation depends on the identity of her prior firm in her first year at each firm. That is, suppose a worker moves from firm 2 to firm 1, then $\mathbb{E}[\iota_1 | V_1^e + \iota_1 > V_2^e + \iota_2] = \mathbb{E}[\iota_1 | \iota_1 - \iota_2 > V_2^e - V_1^e]$. In the second and subsequent years, this selection term for a worker at employer j is

$$(2) \quad \mathbb{E}[\iota | V_j^e, \text{not move}] = \frac{\sum_{\mathcal{E} \backslash j, n} Pr(\text{offer from k and not move}) \mathbb{E}[\iota | \text{offer from k and not move}]}{\sum_{\mathcal{E} \backslash j, n} Pr(\text{offer from k and not move})}.$$

For a worker at j, these terms—when involving other firms—are:

$$(3) \qquad Pr(\text{offer from k and not move}) = \lambda_1 f_k \frac{exp(V_j^e)}{exp(V_k^e) + exp(V_j^e)},$$

$$(4) \qquad \mathbb{E}[\iota | \text{offer from k and not move}] = \gamma - \log\left(\frac{exp(V_j^e)}{exp(V_k^e) + exp(V_j^e)}\right).$$

For a worker at j, these terms are (when involving nonemployment):

$$(5) \qquad Pr(\text{offer from nonemp and not move}) = (1 - \lambda_1) \frac{exp(V_j^e)}{exp(V^n) + exp(V_j^e)},$$

$$(6) \qquad \mathbb{E}[\iota | \text{offer from nonemp and not move}] = \gamma - \log\left(\frac{exp(V_j^e)}{exp(V^n) + exp(V_j^e)}\right).$$

In implementation there are a couple issues. First, for the first year that a worker appears in the dataset I do not know which selection correction term to apply; that is, it might be that the worker showed up from another firm, or it might be that the worker had already been there. To address this, I assume that all such observations are in the second or subsequent years of the employment relationship. Second, there are firms that I cannot estimate the revealed value of, even though I can estimate the value of the firm in the earnings equation (these are firms in the strongly connected set for which I cannot estimate either f or g). For the purposes of the selection correction, I assume that $\frac{g}{f} = 1$ and therefore use the mobility relevant value. Third, to speed up computation, I discretize the firms into $1,000$ equally-sized (in terms of person-years) bins and use the bin means to compute the selection correction.

I then use the expectation of the ι in the earnings equation. Let $\mathbb{E}[\iota | i, t]$ denote the expectation of the ι given the worker i's history. Then I estimate:

$$(7) \qquad y_{it} = \alpha_i + \Psi_{\boldsymbol{J}(i,t)} + x_{it}' \beta + \eta \mathbb{E}[\iota | i, t] + \tilde{r}_{it},$$

where $\tilde{r}$ indicates that this is a different residual than in equation (**??**) because I include the expectation of ι. I find that the correlation of Ψ_j with and without the selection correlation is 0.99986. Table A8 presents the variance decomposition with the selection correction.

D Appendix: Overidentification test

The model has a theory of every entry in the M^o matrix and thus it is possible to construct a wide variety of overidentifying tests by comparing the model predictions for entries to the empirical matrix. The overidentifying test I conduct focuses on the "dense" part of the matrix and asks how well the top eigenvector is able to predict binary comparisons. Specifically, I study the (j, k) pairs where $M_{jk}^o \neq 0$ and $M_{kj}^o \neq 0$. For each such pair, I label the winner of the binary comparison—or the "local" winner— the firm that has the most flows; i.e. j wins if $M_{jk}^o > M_{kj}^o$ and k wins if $M_{kj}^o > M_{jk}^o$. In contrast, the model says that firm k wins if $\tilde{V}_k^o > \tilde{V}_j^o$, or the "global" winner.[10]

The extent of disagreement between the global and local rankings is consistent with the model being the data-generating process. When I weight the comparisons by the number of accepted offers represented in each comparison,[11] the model and the binary comparisons agree on 71.0% of comparisons. Is 70% big or small? This number allows me to reject the null of the model being equivalent to all firms having the same value.[12] I find that the 90% confidence interval under the random null is [49.77%, 50.23%]. Under the null that the model is the data-generating process, the 90% confidence interval is [77.37%, 77.49%].[13] This means that the data are statistically inconsistent with the model being the data-generating process, but the economic magnitude of the rejection is not large. Thus, I conclude that the top eigenvector of the mobility matrix does a reasonable job of summarizing the structure of the employer-to-employer transitions.

E Appendix: Omitted Proofs

Proof of Result ??

Notational/definitional preliminaries: This follows the presentation in Minc (1988) of standard graph theory definitions. Let M be a matrix, where entry M_{ij} measures flows from employer j to employer i. Note that all entries in M are by construction nonnegative: the entries are either zeros, or positive values. Let $\mathcal{E}$ be a set (of employers) labelled from $1...n$. Let A be a set of ordered pairs of elements of $\mathcal{E}$. The pair $D = (\mathcal{E}, A)$ is a *directed* graph. $\mathcal{E}$ is the set of *vertices*, and the elements of A are the *arcs* of D, which represent *directed* flows between employers. A sequence of arcs $(i, t_1)(t_1, t_2)...(t_{m-2}, t_{m-1})(t_{m-1}, j)$ is a *path connecting j to i*. The *adjacency* matrix of a directed graph is the $(0, 1)$ matrix whose (i, j) entry is 1 if and only if (i, j) is an arc of D. An adjacency matrix is associated with a nonnegative matrix M if it has the same zero pattern as M. The directed graph is *strongly connected* if for any pair of distinct vertices i and j there is a path in D connecting i to j and j to i. The directed graph is *connected* if for any pair of distinct vertices i and j there is a path in D connecting i to j or a path connecting j to i.

Proof Observe that if M is strongly connected, then every column sum is nonzero so that the adjacency matrix associated with M is the same as the adjacency matrix associated with $S^{-1}M$.

---

[10]The reason to focus on the top eigenvector is that adjustments for the offer distribution will affect both comparisons in the same way.

[11]I.e., for a comparison of j and k, I weight by $M_{jk}^o + M_{kj}^o$.

[12]If in the data I observe M_{BA}^o workers flowing from A to B and M_{AB}^o workers flowing from B to A, then I take $M_{AB}^o + M_{BA}^o$ draws from a binomial distribution, where the probability of choosing A is 0.5. I ask what share of weighted comparisons the model and the binary comparisons agree on. I repeat this procedure 50 times to generate a null distribution under the hypothesis of all firms are equally appealing.

[13]I repeat the procedure described in footnote 12 except that the probability of choosing A is given by $\frac{exp(\tilde{V}_A)}{exp(\tilde{V}_A)+exp(\tilde{V}_B)}$, where $exp(\tilde{V}_A)$ is what I estimate in the model (and similarly for B) and the probability of choosing B is the remaining probability.

By Minc (1988), chapter 4, theorem 3.2, a nonnegative matrix is irreducible if and only if the associated directed graph is strongly connected. By Minc (1988), chapter 1, theorem 4.4, an irreducible matrix has exactly one eigenvector in E^n (the simplex). If M represents a set of strongly connected firms then these two theorems (often jointly called the Perron-Frobenius theorem) guarantee the existence of a unique solution of the form:

$$S^{-1}Mexp(\tilde{V}) = \lambda exp(\tilde{V}),$$

where all the entries in $exp(\tilde{V})$ are of the same sign.

All that remains to show is that $\lambda = 1$. Consider the jth row of $S^{-1}Mexp(\tilde{V}) = \lambda exp(\tilde{V})$. Let e_j be the basis vector; that is, it is a zero vector with 1 in the jth row.

(8) $$[S^{-1}Mexp(\tilde{V})]_j = [\lambda exp(\tilde{V})]_j,$$

(9) $$\frac{e_j^T Mexp(\tilde{V})}{||Me_j||_1} = \lambda e_j^T exp(\tilde{V}),$$

where $|| \cdot ||_1$ is the l_1 norm of a matrix so for an arbitrary matrix A we have $||A||_1 = \sum_k \sum_j |a_{kj}|$. Note that $||Me_j||_1$ is a scalar.

Because M is a nonnegative matrix, we can rewrite the l_1 norm as a dot product with a vector of ones. Specifically, let $\mathbf{1}$ be a column vector of 1s:

(10) $$||Me_j||_1 = \mathbf{1}^T Me_j.$$

Rearrange:

(11) $$\frac{e_j^T Mexp(\tilde{V})}{||Me_j||_1} = \lambda e_j^T exp(\tilde{V}),$$

(12) $$\frac{e_j^T Mexp(\tilde{V})}{\mathbf{1}^T Me_j} = \lambda e_j^T exp(\tilde{V}),$$

(13) $$e_j^T Mexp(\tilde{V}) = \lambda \mathbf{1}^T Me_j e_j^T exp(\tilde{V}).$$

Now sum over the rows:

(14) $$\sum_j e_j^T Mexp(\tilde{V}) = \sum_j \lambda \mathbf{1}^T Me_j e_j^T exp(\tilde{V}),$$

(15) $$\sum_j e_j^T Mexp(\tilde{V}) = \lambda \sum_j \mathbf{1}^T Me_j e_j^T exp(\tilde{V}),$$

(16) $$\mathbf{1}^T Mexp(\tilde{V}) = \lambda \sum_j \mathbf{1}^T Me_j e_j^T exp(\tilde{V}),$$

(17) $$\mathbf{1}^T Mexp(\tilde{V}) = \lambda \mathbf{1}^T M \sum_j e_j e_j^T exp(\tilde{V}),$$

(18) $$\mathbf{1}^T Mexp(\tilde{V}) = \lambda \mathbf{1}^T Mexp(\tilde{V}).$$

Hence, $\lambda = 1$.

Proof of Result ??

Proof The proof shows that the diagonal elements cancel out. First, use the identity from (**??**):

$$exp(\tilde{V}_j) \sum_{k' \in \mathcal{E} \cup n} M_{k'j} = \sum_{k \in \mathcal{E} \cup n} M_{jk} exp(\tilde{V}_k).$$

Expand to write the diagonal elements explicitly:

$$exp(\tilde{V}_j) M_{jj} + exp(\tilde{V}_j) \sum_{k' \in \mathcal{E} \cup n \backslash \{j\}} M_{k'j} = \sum_{k \in \mathcal{E} \cup n \backslash \{j\}} M_{jk} exp(\tilde{V}_k) + exp(\tilde{V}_j) M_{jj}.$$

Then cancel the diagonal terms to show that (**??**) holds with arbitrary diagonal elements:

$$exp(\tilde{V}_j) \sum_{k' \in \mathcal{E} \cup n \backslash \{j\}} M_{k'j} = \sum_{k \in \mathcal{E} \cup n \backslash \{j\}} M_{kj} exp(\tilde{V}_k).$$

F   Appendix: Alternative derivation of the decomposition in section ??

Result 1 *Suppose that the utility function is given by equation (**??**) and that $\{V_j^e\}_{j \in \mathcal{E}}$ and $\{\Psi_j\}_{j \in \mathcal{E}}$ are known. Then*

$$Var(a_{Rosen}) = (1 - R^2) Var(\Psi),$$
$$Var(a_{Mortensen}) \in [0, \infty)$$

and combining them, we have

$$Var(a) \in [Var(\Psi)(1 - R^2), +\infty),$$

where $R^2 = Corr(V^e, \Psi)^2$. The willingness to pay for Rosen *and* Mortensen *amenities is one. The* Rosen *amenities are related to earnings as follows:*

$$Corr(\Psi, a_{Rosen}) = -\sqrt{1 - R^2}.$$

When $Var(a_{Mortensen}) > 0$,

$$Corr(\Psi, a_{Mortensen}) = \sqrt{R^2}.$$

Bounds on the variance of utility in log dollar units are:

$$Var(\Psi + a) \in [Var(\Psi) R^2, \infty).$$

Proof It is helpful to first have explicit expressions for a number of quantities. Write the R^2 between Ψ and V in terms of the known variable V and the unknown variable a:

$$(19) \qquad R^2 = \frac{Cov(\Psi, V)^2}{Var(\Psi)Var(V)}$$

$$(20) \qquad = \frac{Cov(\Psi, \alpha(\Psi + a))^2}{Var(\Psi)Var(\alpha(\Psi + a))}$$

$$(21) \qquad = \frac{\alpha^2 Cov(\Psi, (\Psi + a))^2}{\alpha^2 Var(\Psi)Var((\Psi + a))}$$

$$(22) \qquad = \frac{[Var(\Psi) + Cov(\Psi, a)]^2}{Var(\Psi)[Var(\Psi) + Var(a) + 2Cov(\Psi, a)]}.$$

It is also helpful to write $Var(a)$ in terms of one unknown quantity by rearranging equation (22):

(23)

$$R^2[Var(\Psi)^2 + Var(\Psi)Var(a) + 2Var(\Psi)Cov(\Psi, a)] = Var(\Psi)^2 + 2Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2,$$

$$(24) \qquad R^2 Var(\Psi)Var(a) = (1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2,$$

$$(25) \qquad Var(a) = \frac{(1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2}{R^2 Var(\Psi)}.$$

The following is a useful expression for $Corr(\Psi, a)$:

$$(26) \qquad Corr(\Psi, a) = \frac{Cov(\Psi, a)}{\sqrt{Var(a)Var(\Psi)}},$$

$$(27) \qquad = \frac{Cov(\Psi, a)}{\sqrt{\frac{(1-R^2)Var(\Psi)^2 + 2(1-R^2)Var(\Psi)Cov(\Psi,a) + Cov(\Psi,a)^2}{R^2 Var(\Psi)} Var(\Psi)}},$$

$$(28) \qquad = \sqrt{R^2} \frac{Cov(\Psi, a)}{\sqrt{(1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2}}.$$

A lower bound on $Var(a)$: To minimize $Var(a)$, start with the expression for $Var(a)$ (equation (25)) in terms of $Cov(\Psi, a)$ and take the first order condition with respect to $Cov(\Psi, a)$:

$$(29) \qquad \frac{\partial Var(a)}{\partial Cov(\Psi, a)} = \frac{2(1 - R^2)Var(\Psi) + 2Cov(\Psi, a)}{R^2 Var(\Psi)}$$

$$(30) \qquad 0 = \frac{2(1 - R^2)Var(\Psi) + 2Cov(\Psi, a)}{R^2 Var(\Psi)}$$

$$(31) \qquad Cov(\Psi, a) = -(1 - R^2)Var(\Psi).$$

The second order condition is $\frac{2}{R^2 Var(\Psi)}$, which is positive. Substitute this into the expression for $Var(a)$ (equation (25)) to get that the minimum value is given by:

$$(32) \quad Var(a) = \frac{(1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)(-(1 - R^2)Var(\Psi)) + (-(1 - R^2)Var(\Psi))^2}{R^2 Var(\Psi)}$$

$$(33) \qquad = Var(\Psi)(1 - R^2).$$

Compute the correlation between Ψ and a at this lower bound:

$$(34) \qquad Corr(\Psi, a) = \frac{Cov(\Psi, a)}{\sqrt{Var(a)Var(\Psi)}}$$

$$(35) \qquad = \frac{-(1 - R^2)Var(\Psi)}{\sqrt{Var(\Psi)(1 - R^2)Var(\Psi)}}$$

$$(36) \qquad = -\sqrt{1 - R^2}.$$

And compute the variance of utility in log dollar units:

$$(37) \qquad Var(\Psi + a) = Var(\Psi) + Var(a) + 2Cov(\Psi, a)$$

$$(38) \qquad = Var(\Psi) + Var(\Psi)(1 - R^2) - 2(1 - R^2)Var(\Psi)$$

$$(39) \qquad = R^2 Var(\Psi).$$

An upper bound on $Var(a)$: Take the limit of the expression for $Var(a)$ (equation (25)) while treating R^2 as a constant (because it is observable data):

$$(40) \qquad \lim_{Cov(\Psi, a) \to \infty} \frac{(1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2}{R^2 Var(\Psi)} = \infty.$$

Note that this implies that $Var(a)$ goes to infinity with the square of $Cov(\Psi, a)$, which is why the R^2 expression remains finite.

What is $Corr(\Psi, a)$ in this case?

$$(41)$$
$$\lim_{Cov(\Psi, a) \to \infty} Corr(\Psi, a) = \lim_{Cov(\Psi, a) \to \infty} \sqrt{R^2} \frac{Cov(\Psi, a)}{\sqrt{(1 - R^2)Var(\Psi)^2 + 2(1 - R^2)Var(\Psi)Cov(\Psi, a) + Cov(\Psi, a)^2}}$$

$$(42) \qquad = \sqrt{R^2}$$

$$(43) \qquad = Corr(\Psi, V).$$

And:

$$(44) \qquad Var(\Psi + a) \to \infty.$$

Rosen vs. Mortensen amenities: To decompose the a term into *Rosen* and *Mortensen* amenities, note that the properties of the *Rosen* amenities correspond to the lower bounds in these results, while the properties of the *Mortensen* amenities correspond to the upper bounds.

To see that the a_{Rosen} term captures variation in pay while holding value constant, consider the following equation:

$$(45) \qquad \Psi = \beta V^e + \epsilon.$$

Treating this equation as a regression (where we have demeaned V^e and Ψ, which is without loss of generality because they are only identified up to location), we have:

$$(46) \qquad \hat{\beta} = \frac{Cov(\Psi, V^e)}{Var(V^e)}.$$

And:

$$(47) \qquad \hat{\epsilon} = \Psi - \hat{\beta} V^e.$$

Note that $\hat{\epsilon} = -a_{Rosen}$ because it generates variation in pay (Ψ) while holding value constant. Hence, a_{Rosen} and Ψ are in the same units and so workers are willing to trade them off one-for-one.

For the variance of a_{Rosen}:

$$(48) \qquad Var(a_{Rosen}) = Var(\hat{\beta} V^e - \Psi)$$

$$(49) \qquad = Var(\Psi) + \hat{\beta}^2 Var(V^e) - 2\hat{\beta} Cov(\Psi, V^e)$$

$$(50) \qquad = Var(\Psi) + \left(\frac{Cov(\Psi, V^e)}{Var(V^e)} \right)^2 Var(V^e) - 2 \left(\frac{Cov(\Psi, V^e)}{Var(V^e)} \right) Cov(\Psi, V^e)$$

$$(51) \qquad = Var(\Psi) - \frac{Cov(\Psi, V^e)^2}{Var(V^e)}$$

$$(52) \qquad = Var(\Psi) - \frac{Cov(\Psi, V^e)^2}{Var(V^e) Var(\Psi)} Var(\Psi)$$

$$(53) \qquad = Var(\Psi)(1 - R^2).$$

For the covariance of a_{Rosen} and Ψ:

$$(54) \qquad Cov(a_{Rosen}, \Psi) = Cov(\hat{\beta} V^e - \Psi, \Psi)$$

$$(55) \qquad = \hat{\beta} Cov(V^e, \Psi) - Var(\Psi)$$

$$(56) \qquad = \left(\frac{Cov(\Psi, V^e)}{Var(V^e)} \right) Cov(V^e, \Psi) - Var(\Psi)$$

$$(57) \qquad = \frac{Cov(\Psi, V^e)^2}{Var(V^e) Var(\Psi)} Var(\Psi) - Var(\Psi)$$

$$(58) \qquad = -Var(\Psi)(1 - R^2).$$

Finally, for the correlation of a_{Rosen} and Ψ:

$$(59) \qquad Corr(a_{Rosen}, \Psi) = \frac{Cov(a_{Rosen}, \Psi)}{\sqrt{Var(a_{Rosen}) Var(\Psi)}}$$

$$(60) \qquad = \frac{Var(\Psi)(1 - R^2)}{\sqrt{Var(\Psi)(1 - R^2) Var(\Psi)}}$$

$$(61) \qquad = -\sqrt{1 - R^2}.$$

These are exactly the properties of a at the lower bounds.

The properties of a at the upper bound correspond to the properties of $a_{Mortensen}$ (conditional on the variance being positive). In terms of interpretation, $Corr(\Psi, a_{Mortensen}) > 0$ means that a hedonic regression would find a wrong-signed coefficient on $a_{Mortensen}$ and hence corresponds to the explanation for the absence of evidence of compensating differentials that desirable nonpay characteristics are positively correlated with pay. Note that $a_{Mortensen}$ does not correspond to nonpay characteristics that are orthogonal to pay.

Willingness to pay for Rosen and Mortensen amenities: Note that in these derivations $V^e = \omega(\Psi + a)$ so that by construction the Rosen and Mortensen amenities are in the same units as Ψ (log dollars) and workers are willing to trade-off one-for-one between log dollars and the amenities.

G Appendix: Description of estimating the model

Step 0 Initialize the model EE and EN probabilities using EE and EN probabilities at expanding firms. The relative size of employers (g_j) and the number of workers (W) are summary statistics of the data. Initalize $\{V_j^e\}_{j \in \mathcal{E}}$ to be a constant.

Step 1 *Exogenous separations*: Using the method described in section **??**, build M and compute δ and ρ using these probabilities.

Step 2 *Central tendency of worker flows*: Using equation (**??**), compute $exp(\tilde{V})$.

Step 3 *Offer distribution*: Compute f by doing a grid-search on λ_1 to match the level of EE flows. As an output this gives a new value of $\{V_j^e, f_j\}_{j \in \mathcal{E}}$ as well as λ_1. See below for more detail.

Step 4 Given the new values of $\{V_j^e\}_{j \in \mathcal{E}}$ use equation (**??**) to compute the new counterfactual separation probabilities. If the size-weighted correlation between the old and new $\{V_j^e\}_{j \in \mathcal{E}}$ is less than 0.999, then return to step 1.

Details on step 3

Define C_1 to be the share of offers that are accepted from nonemployment, or:

$$(62) \qquad C_1 \equiv \sum_{j' \in \mathcal{E}} f_{j'} \frac{exp(V_j^e)}{exp(V_{j'}^e) + exp(V^n)},$$

so that f_j^o can be written in terms of model parameters as

$$(63) \qquad f_j^o = \frac{f_j \frac{exp(V_j^e)}{exp(V_j^e) + exp(V^n)}}{C_1}.$$

Take an initial guess of λ_1:

- Evaluate two equations, where I maintain the convention of data or variables whose values are known by a given step are on the left-hand side, while unknowns are on the right-hand side. In the following equation, g_j and f_j^o are from step 0, δ_j and ρ_j are from step 1, and $\tilde{V}_j$ is from step 2. The first equation is an identity, where the right hand side comes from substituting in equations (**??**) and (63) to the left hand side:

$$(64)$$
$$\frac{g_j exp(\tilde{V}_j)}{f_j^o}(1 - \delta_j)(1 - \rho_j) = g_j \frac{f_j exp(V_j^e)}{g_j(1 - \delta_j)(1 - \rho_j)} C_1 \frac{1}{f_j} \frac{exp(V_j^e) + exp(V^n)}{exp(V_j^e)}(1 - \delta_j)(1 - \rho_j)$$
$$(65) \qquad\qquad\qquad = C_1[exp(V_j^e) + exp(V^n)].$$

The second equation comes from rewriting equation (??) using the C_1 notation:

(66)
$$\frac{1}{1-\lambda_1}\frac{1}{W}\sum_{j\in\mathcal{E}}M_{jn}exp(\tilde{V}_n) = \frac{1}{1-\lambda_1}\frac{1}{W}\sum_{j\in\mathcal{E}}\lambda_0 U f_j \frac{exp(V_j^e)}{exp(V^n)+exp(V_j^e)}\frac{(1-\lambda_1)W exp(V^n)}{\lambda_0 U}$$

(67)
$$= \frac{1}{1-\lambda_1}(1-\lambda_1)exp(V^n)\sum_{j\in\mathcal{E}}f_j\frac{exp(V_j^e)}{exp(V^n)+exp(V_j^e)}$$

(68)
$$= exp(V^n)C_1.$$

- Combine equations (65) and (68), to give the following two terms: $C_1 exp(V_j^e)$ and $C_1 exp(V^n)$.

- Rewrite equation (63) by multiplying by $\frac{C_1}{C_1}$ and rearranging:

(69)
$$f_j^o = f_j\frac{exp(V_j^e)}{exp(V_j^e)+exp(V^n)}\frac{1}{C_1}$$

(70)
$$= f_j\frac{C_1 exp(V_j^e)}{C_1 exp(V_j^e)+C_1 exp(V^n)}\frac{1}{C_1}$$

(71)
$$f_j^o\frac{C_1 exp(V_j^e)+C_1 exp(V^n)}{C_1 exp(V_j^e)} = \frac{f_j}{C_1}.$$

In this equation, the terms on the left-hand side are known from step 1, so this step gives $\frac{f_j}{C_1}$.

- Now that $\frac{f_j}{C_1}$ is known, solve for C_1 by using the normalization $\sum_{j\in\mathcal{E}}f_j = 1$ (note that C_1 contains f_j, so the scale transformation cancels out.[14]).

(72)
$$\sum_{j\in\mathcal{E}}\frac{f_j}{C_1} = \frac{\sum_{j\in\mathcal{E}}f_j}{C_1} = \frac{1}{C_1}.$$

Now that C_1 is known and from equation (71) $\frac{f_j}{C_1}$ is known, it is possible to solve for f_j.

- Knowledge of C_1 gives $exp(V^n)$ and $exp(V_j^e)$, via equations (65) and (68).

- Given the parameters of the model, compute the number of endogenous employer-to-employer transitions implied by the model:[15]

(73)
$$\lambda_1\sum_i g_j(1-\delta_j)(1-\rho_j)\sum_k f_k\frac{exp(V_k^e)}{exp(V_k^e)+exp(V_j^e)}.$$

I search over a grid of width 0.001 and select the λ_1 that minimizes the absolute gap between equation (73) and the probability of EE transitions in the data, or

(74)
$$\frac{\sum_{j\in\mathcal{E}\setminus\{j\}}\sum_{k\in\mathcal{E}}M_{jk}}{W\sum_{j\in\mathcal{E}}g_j(1-\delta_j)(1-\rho_j)}.$$

---

[14] Define $\hat{f}_j = \alpha f_j$ and let $\hat{C}_1$ be the C_1 constructed using $\hat{f}_j$. Then $\frac{\hat{f}_j}{\hat{C}_1} = \frac{f_j}{C_1} = \frac{\alpha f_j}{\sum_{j'\in\mathcal{E}}\alpha f_{j'}\frac{exp(V_{j'}^e)}{exp(V_{j'}^e)+exp(V^n)}}$

[15]To make this computationally feasible, group firms into $1,000$ categories on the basis of the firm values (V^e).

H Appendix: Addressing measurement error

This appendix provides more details on two approaches to addressing and quantifying the role of measurement error in driving my results.

Approach 1: Shrinkage

The first approach uses an empirical Bayes approach to shrinkage. Specifically, I follow Morris (1983) and use estimates of the standard errors to downweight noisier observations. In my context, measurement error means that I overstate the variance of the underlying values and so understate the correlation.

Standard errors

To compute standard errors, I use the bootstrap.

To maintain the dependency structure in the data, I resample at the level of worker-year-pairs. That is, if I have three earnings observations for worker w, $\{y_{i,t-1}, y_{i,t}, y_{i,t+1}\}$, then I create a set of two observations, $\{(y_{i,t-1}, y_{i,t}), (y_{i,t}, y_{i,t+1})\}$, where I record how the worker moved (i.e., EE, ENE, or not at all) from the employer in the first period to the second period.[16] The asymptotic thought experiment that this relates to is allowing W in equation (??) to grow.

Two issues arise in the bootstrap: first, how to normalize estimates and second, the identified set. Because both V_j^e and Ψ_j are only identified up to location, I need to normalize the value of one j. I normalize the location of Ψ by assuming that the estimates are noiseless for a very large firm. I normalize the location of V_j^e by setting $V^n = 0$ in all repetitions.

The second issue is that the identified set of firms varies across bootstrap resamples. The reason is that the strongly connected set of firms differs across the bootstrap resamples. This means that there will be a different number of observations to estimate the variability of each of the parameters (that is, smaller firms will typically have fewer resamples). I address this in two steps. First, my initial sample selection eliminates the smallest firms where this issue is likely to come up the most (i.e., I eliminate firms that have fewer than 90 non-singleton observations). Second, I compute 50 bootstrap replications and only keep firms that show up in at least 20 of them.

Shrinkage

I use the standard errors to shrink the estimates of V_j^e and Ψ_j. Formally, I follow the empirical Bayes approach laid out by Morris (1983). My exposition follows Online Appendix C in Chandra et al. (2016).

Define some notation. Let j be a firm. Let n_J be the number of firms. Let $n_{p(j)}$ be the number of person years represented by firm j. Let q_j be a measure of the quality of the firm—i.e., either V_j^e or Ψ_j. Let $\hat{q}_j$ denote the estimate of q for firm j. Let Q be the $n_j \times 1$ vector of $\hat{q}_j$. Let $\hat{\pi}_j^2$ denote the variance of the estimate. Let $\hat{\sigma}^2$ denote the estimate of the true variance of q_j. Let $\mathbf{x}_j$ be an $n_x \times 1$ vector of characteristics of firm j. I use a set of dummies for 4-digit industry and county. Let X be the stacked vector of the $\mathbf{x}_j'$. Let $\boldsymbol{\lambda}$ be an $n_x \times 1$ vector of coefficients. Finally, let w_j be the weight of firm j and W be the $n_J \times n_J$ matrix with w_j on the diagonal.

---

[16]While this procedure places equal weight on each transition between employers, it double-counts the interior earnings observations (in this case, the year t earnings observations). Hence, to compute the earnings decomposition in the bootstrap resamples I delete duplicate interior earnings observations. Formally, if an interior earnings observation appears n times in the bootstrap replicate, then I include the observation $\lceil \frac{n}{2} \rceil$ times, where this notation is the ceiling operator that rounds up to the nearest integer.

The following equations show how these terms relate.

$$(75) \qquad w_j = n_{p(j)} \frac{1}{\hat{\pi}_i^2 + \hat{\sigma}^2}$$

$$(76) \qquad \hat{\sigma}^2 = \max\left\{0, \frac{\sum_j w_j\left\{\frac{n_J}{n_J - n_x}(\hat{q}_j - \boldsymbol{x}_j'\boldsymbol{\lambda})^2 - \hat{\pi}_j^2\right\}}{\sum_j w_j}\right\}$$

$$(77) \qquad \hat{\boldsymbol{\lambda}} \equiv (X'WX)^{-1}X'WQ.$$

The two unknowns are $\hat{\sigma}^2$ and $\hat{\boldsymbol{\lambda}}$. These are solved for in a loop. Initialize $w_j = n_{p(j)}$. Then iterate the following till convergence:

1. Compute $\hat{\boldsymbol{\lambda}}$, then a new estimate of $\hat{\sigma}^2$ (using the above equation).

2. Check if $\hat{\sigma}^2$ has converged. If not, update the weights, w_j, and return to step 1.

The feasible shrinkage estimator is:

$$(78) \qquad \hat{b}_j = \left(\frac{n_J - n_x - 2}{n_J - n_x}\right)\left(\frac{\hat{\pi}_j^2}{\hat{\pi}_j^2 + \hat{\sigma}^2}\right)$$

$$(79) \qquad q_j^{EB(f)} = (1 - \hat{b}_j)\hat{q}_j + \hat{b}_j\boldsymbol{x}_j'\hat{\boldsymbol{\lambda}}.$$

The variance of the distribution unconditional on covariates is given by:

$$(80) \qquad \hat{\zeta}^2 = \max\left\{0, \frac{\sum_j w_j\left\{\frac{n_J}{n_J - 1}(\hat{q}_j - \bar{q})^2 - \hat{\pi}_j^2\right\}}{\sum_j w_j}\right\},$$

where

$$(81) \qquad \bar{q} = \frac{\sum_j w_j \hat{q}_j}{\sum_j w_j}.$$

Now suppose we have two measure of firm quality A and B and we want to know their correlation. Let a tilde'd variable represent a variable that is adjusted for measurement error. Then:

$$(82) \qquad \widetilde{Corr}_j(q_A, q_B) = \frac{Cov(\hat{q}_A, \hat{q}_B)}{\sqrt{\hat{\zeta}_A^2 \hat{\zeta}_B^2}},$$

$$(83) \qquad \widetilde{R}^2(q_A, q_B) = \left(\widetilde{Corr}_i(q_A, q_B)\right)^2,$$

where this reflects the assumption that the measurement error in A and B is uncorrelated.

Recall that $\hat{\boldsymbol{\lambda}}$ is the vector of coefficients, which reflects industry and location means of the measure of quality. For some purposes I am interested in comparing these. Hence, I shrink $\hat{\boldsymbol{\lambda}}$ using the observation that $\hat{\boldsymbol{\lambda}}$ is computed from the following regression:

$$(84) \qquad \sqrt{W}Q = \boldsymbol{\lambda}\sqrt{W}X,$$

and so I can estimate the variance around the $\hat{\boldsymbol{\lambda}}$ using analytical formulas for the variance-covariance matrix, and then shrink these estimates using the formulas above.

Approach 2: Split samples

A second approach to quantifying the importance of measurement error is to split the sample in half. Specifically, I divide the sample based on *people* (and unconditional on firm) by randomly allocating each unique person into sample 1 or sample 2. By splitting on the basis of people, I get two independent estimates of the value and pay at each firm that shows up in strongly connected set defined by each subsample. With two independent estimates of the same quantity, I can estimate how much of the variance is due to noise.

Formally, let $\hat{V}_j^{e,1}$ be the estimate in subsample 1 and $\hat{V}_j^{e,2}$ be the estimate of the value of being employed at firm j in subsample 2. Assume that:

$$(85) \qquad \hat{V}_j^{e,1} = V_j^e + \epsilon_{j,1}$$

and

$$(86) \qquad \hat{V}_j^{e,2} = V_j^e + \epsilon_{j,2}.$$

Because the samples are mutually exclusive, the errors are uncorrelated and $Cov(\epsilon_{j,1}, \epsilon_{j,2}) = 0, Cov(\hat{V}_j^{e,2}, \epsilon_{j,1}) = 0$, and $Cov(\hat{V}_j^{e,1}, \epsilon_{j,2}) = 0 \ \forall \ j$. Hence,

$$(87) \qquad Corr(\hat{V}_j^{e,1}, \hat{V}_j^{e,2}) = \frac{Cov(\hat{V}_j^{e,1}, \hat{V}_j^{e,2})}{\sqrt{Var(\hat{V}_j^{e,1})Var(\hat{V}_j^{e,2})}}$$

$$(88) \qquad = \frac{Cov(V_j^e, V_j^e) + 2Cov(\epsilon_{j,1}, \epsilon_{j,2})}{\sqrt{Var(\hat{V}_j^{e,1})Var(\hat{V}_j^{e,2})}}$$

$$(89) \qquad = \frac{Var(V_j^e)}{\sqrt{Var(\hat{V}_j^{e,1})Var(\hat{V}_j^{e,2})}}.$$

The core exercise of this paper reduces to:

$$(90) \qquad R^2(\hat{V}_j^e, \hat{\Psi}_j^e) = \frac{Cov(\hat{V}_j^e, \hat{\Psi}_j^e)}{Var(\hat{V}_j^e)Var(\hat{\Psi}_j^e)}.$$

Under the assumption that $Cov(\hat{V}_j^e, \hat{\Psi}_j^e) = Cov(V_j^e, \Psi_j^e)$, we have:

$$(91) \qquad R^2(V_j^e, \Psi_j^e) = \frac{Cov(V_j^e, \Psi_j^e)^2}{Var(V_j^e)Var(\Psi_j^e)}$$

$$(92) \qquad < \frac{R^2(\hat{V}_j^e, \hat{\Psi}_j^e)}{Corr(\hat{V}_j^{e,1}, \hat{V}_j^{e,2})Corr(\hat{\Psi}_j^{e,1}, \hat{\Psi}_j^{e,2})}.$$

I   Appendix: Monte Carlo evidence

This appendix describes Monte Carlo evidence on the properties of the estimators used in this paper.

Notation reminder

To make this appendix self-contained, let me remind the reader of some notation (all of these values are estimated, but for notational simplicity I omit notation to capture this):

- δ_j is the exogenous EN separation rate at j;

- ρ_j is the exogenous EE separation rate at j;

- g_j is the share of employment at j;

- W is the total number of non-singleton person-years in the data (i.e., years where I see the worker again);

- f_j is the share of offers at j;

- λ_1 is the probably of getting an EE offer;

- Ψ_j is the pay at firm j;

- V_j^e is the value of being employed at j;

- V^n is the value of nonemployment.

Simulation details

There are a few high-level issues that inform the design of the Monte Carlos. First, the search model does not impose steady state so that firms can grow or shrink over time. Hence, if I take the set of parameter values and run the model for a large number of periods then, in the limit, the data would be dominated by a small number of firms. Second, the method to identify exogenous separations relies on variation in the growth rates of a given firm over time. Hence, in order to have a simulation that generates data where it is possible to include all steps of estimation, I need to have multiple time periods. Third, as is well-known, in order to run AKM it is necessary to have multiple observations per worker and to follow workers across firms (whereas in the search model the identify of workers is irrelevant). Fourth, it is computationally intensive to estimate the model.

Narrative: To balance these various considerations, I proceed as follows. First, to address the fact that it is computationally intensive to estimate the model, I randomly sample of firms. To preserve the size distribution of firms, I stratify the firms by size and then sample from each bin, where the bins contain an equal number of person-years. The bootstrapping is especially time-intensive, so I report a one-in-200 sample for the bootstrapping results and a one-in-10 sample for all results.[17] Second, to generate multiple observations over time, I divide the number of non-singleton person-years at a firm in four ($\frac{g_j W}{4}$) and thus allow for five periods. Third, to allow for variation in firm growth rates, I set the realization of δ_j and ρ_j to zero in three periods, and equal to $4\{\delta_j, \rho_j\}$ in one period (I do not tell the model estimation code which time period this occurs). Fourth, to generate a panel of workers, in each of the four years I follow workers between the two employers. If the firm grows from one year to the next, then I cut-off worker histories randomly. In contrast, if the firm shrinks, then I add new workers. The net result of this procedure is that workers appear in the simulated data for at most five periods, but there is a distribution of the length of the worker's labor market histories.

---

[17]To show the effect of varying network density, I also report one-in-20 samples.

Algorithmic: In what follows, X refers to the 1-in-X sample, where i report the results for $X \in \{10, 20, 200\}$,:

1. Draw a sample of firms from the core sample of firms in estimation (column (4) of Table **??**):

   - Sort firms based on size (g_j);
   - Divide into X bins, where $\sum_j g_j$ for the j in the bin is the same;
   - Draw a 1-in-X sample of j from each bin; this step gives the set of j that appear in the simulation run;
   - Renormalize the f_j to sum to one in this subsample.

2. Generate 4 periods of data, where there are $\frac{g_j W}{4}$ workers who start each period at firm j.

   - For 3 periods of data, workers make mobility decisions where the probability of an exogenous separation is zero. The worker receives an EE offer with probability λ_1.
     - If the realization is 1, then the offer is drawn from the offer distribution, and the worker's acceptance decision is a Bernoulli random variable where the acceptance probability is given by the model (for an offer from k, the probability of accepting is $\frac{exp(V_k^e)}{exp(V_k^e)+exp(V_j^e)}$). If the realization of this Bernoulli random variable is 1, this generates an EE transition from j to k. Otherwise, it does not generate mobility.
     - If the realization is 0, then the worker always receives an offer from nonemployment. For a worker at j, "quit" decision is a Bernoulli random variable where the quit probability is given by $\frac{exp(V^n)}{exp(V^n)+exp(V_j^e)}$. If the realization of this Bernoulli random variable is 1, this generates two transitions. The first is an EN transition from j. The second transition is an NE transition, where the probability that the worker ends up at k is given by $f_k \frac{exp(V_k^e)}{exp(V_k^e)+exp(V^n)}$.
   - For the 1 period of data where the exogenous shocks are "turned on", then things proceed as follows:
     - First, a Bernoulli random variable is drawn with probability of 1 given by $4\delta_j$. If the draw comes up 1, then this results in two transitions. The first transition is an EN transition from j. The second transition is an NE transition, where the probability that the worker ends up at k is given by $f_k \frac{exp(V_k^e)}{exp(V_k^e)+exp(V^n)}$.
     - Second, if the previous Bernoulli random variable had a realization of 1, then the period is over. If the realization was 0, then a second Bernoulli random variable is drawn where the probability of a 1 is given by $4\rho_j$. If the draw comes up 1, then this results in a single EE transition, where the destination is drawn from f (i.e., it is to k with probability f_k).
     - Third, if the previous Bernoulli random variable had a realization of 1, then the period is over. If the realization was 0, then proceed as in the previous step.

3. For the purposes of estimating the search model, the previous step is sufficient in that it can be used to compute 4 separate M_t^o matrices (including a row and column for nonemployment), and we know the size of the employer in each period and so we can infer the employer growth rate in each period.

4. For the purposes of estimating AKM, assign identities to workers as follows:

   - In the first sub-period of the first period assign each worker a unique identity.

- At the end of the second sub-period in the first period, worker i will either have remained at firm j, or will have moved to firm k. If the total number of workers at a firm in the second sub-period exceeds the number of workers at the beginning of the sub-period, then end a random subset of the worker histories (so that, for the purposes of AKM this generates a worker with a single period of earnings). If the total number of workers at firm j in the second sub-period is the same or fewer than the number of workers at the beginning of the sub-period, then preserve the identity of the worker, and continue this worker to the next period.

- At the beginning of the second period, generate new workers if the number of workers at the firm j that continue from the first period is less than $\frac{g_j W}{4}$.

- Continue as in the second step.

5. For the purposes of AKM, the previous step generates a set of worker-firm pairs, and workers with mobility histories. To generate earnings:

- The Ψ_j are drawn from the data. Denote by σ_Ψ^2 the variance of the Ψ.
- Draw the α_i from a $N(0, \frac{0.57}{0.21}\sigma_\Psi^2)$.
- Draw the ϵ_{it} from a $N(0, \frac{0.57}{0.11}\sigma_\Psi^2)$.
- Draw the covariates, but assign coefficients of zero to them.

I report the results of 100 simulation runs.

Results

Table A10 reports the results of these simulations. The left-hand side of the table shows results from one in 10 sampling, the middle reports results from one in 20 sampling, and the right-hand side of the table shows the results from one in 200 sampling. Panel A reports the (percentage point) gap between the true and estimated values that are relevant for the quantitative bottom-line of the paper, while the italicized rows report the levels of the true values.[18] The basic point to take from Panel A is that the estimation procedure is slightly biased down, but the bias is quantitatively small. Focusing on the one in 10 sampling, the bias down in the raw measure is two percentage points while the bias in the various corrected measures is one percentage point. The basic point to take from comparing the one in 10 sampling and the one in 20 sampling is that the different sampling rate does not have a big effect on the estimates. For example, the median gap between the true and estimated R^2 between V^e and Ψ are both one percentage point. Finally, the one in 200 sampling shows that the bootstrap performs similarly to the split sample approach.

Panel B reports the correlation between the true and estimated values of Ψ, V^e, and $\tilde{V}^{EE}$ across simulation runs. The table allows us to understand the result in Panel A that the procedure is not particularly biased. Specifically, the table shows that given the sample sizes, Ψ and V^e are estimated without that much noise. The correlation between the true and estimated Ψ is 1.00, and the correlation between the true and estimated V^e is also very high, 0.98. In contrast, the table shows that the estimation of the $\tilde{V}^{EE}$ is much noisier. Why is $\tilde{V}^{EE}$ so much noisier than V^e? The basic reason is that estimation of $\tilde{V}^{EE}$ uses much less information than the estimation of V^e: specifically, it only uses information in the EE transitions, whereas V^e also uses information in

---

[18]The levels of the true values vary across simulations because the weights are stochastic. To be specific, even though the set of true V^e and Ψ are fixed across simulation runs, the randomness in which firms end up in the connected set generates variation in firm size because I only keep person-years in which the worker reappears in the dataset.

the ENE transitions (this information is necessary to estimate the value of nonemployment, which is necessary to estimate the offer distribution). Since Table **??** shows that EE transitions account for only 40% of all transitions, it is not surprising that when we move from $\tilde{V}^{EE}$ to V^e that the estimation becomes less noisy.

J Appendix: Other model-consistent approaches to ranking firms

The first approach looks at worker inflows and ranks firms based on the share of hires that are on employer-to-employer transitions. Bagger and Lentz (2017, pg. 21) term this ratio the "poaching index." Formally, the observed poaching index is:

$$(93) \qquad PR_j^o = \frac{\sum_{k \in \mathcal{E} \backslash i} M_{kj}^o}{\sum_{k \in \mathcal{E} \cup n \backslash j} M_{jk}^o}.$$

The model-consistent version focuses on the endogenous flows:

$$(94) \qquad PR_j^m = \frac{\sum_{k \in \mathcal{E} \backslash j} M_{kj}}{\sum_{k \in \mathcal{E} \cup n \backslash j} M_{jk}}.$$

Result 2 *If the value of nonemployment (V^n) is low enough relative to the distribution of the value of employment (V_j^e), then PR_j^m is monotonically increasing in firm value.*

Proof It is easier to work with a monotone transformation of the poaching index and look at the ratio of hires on employer-to-employer transitions, and consider employer-to-employer flows from a firm to itself:

$$(95) \qquad PR_j^m = \frac{\sum_{k \in \mathcal{E}} M_{jk}}{M_{jn}}$$

$$(96) \qquad = \frac{\sum_{k \in \mathcal{E}} g_k W (1-\delta)(1-\rho)\lambda_1 f_j \frac{exp(V_j^e)}{exp(V_j^e)+exp(V_k^e)}}{\lambda_0 U f_j \frac{exp(V_j^e)}{exp(V_j^e)+exp(V^n)}}$$

$$(97) \qquad = \frac{\sum_{k \in \mathcal{E}} g_k W (1-\delta)(1-\rho)\lambda_1 \frac{exp(V_j^e)+exp(V^n)}{exp(V_j^e)+exp(V_k^e)}}{\lambda_0 U}$$

$$(98) \qquad \propto \sum_{k \in \mathcal{E}} g_k \frac{exp(V_j^e) + exp(V^n)}{exp(V_j^e) + exp(V_j^e)},$$

where $\propto$ means "proportional to" and drops all the constant terms. Consider how each term in the sum depends on $exp(V_j^e)$:

$$(99) \qquad \frac{\partial}{\partial exp(V_j^e)} \left[\frac{exp(V_j^e)+exp(V^n)}{exp(V_j^e)+exp(V_k^e)} \right] = \frac{(exp(V_j^e)+exp(V_k^e)) - (exp(V_j^e)+exp(V^n))}{(exp(V_j^e)+exp(V_k^e))^2}$$

$$(100) \qquad = \frac{exp(V_k^e)-exp(V^n)}{(exp(V_j^e)+exp(V_k^e))^2}.$$

Aggregate over all k:

$$(101) \qquad \frac{\partial}{\partial exp(V_j^e)} \left[\sum_{k \in \mathcal{E}} g_k \frac{exp(V_j^e) + exp(V^n)}{exp(V_j^e) + exp(V_k^e)} \right] = \sum_{k \in \mathcal{E}} g_k \frac{exp(V_k^e) - exp(V^n)}{(exp(V_j^e) + exp(V_k^e))^2}.$$

We want conditions on V^n such that this expression is positive for all values of $exp(V_j^e)$ so that PR_j^m is increasing in $exp(V_j^e)$. If $V^n < min_k V_k^e$, then this is always true. Fixing $exp(V_j^e)$, equation (101) is monotone decreasing in $exp(V^n)$:

$$(102) \qquad \frac{\partial}{\partial exp(V^n)} \left[\sum_{k \in \mathcal{E}} g_k \frac{exp(V_k^e) - exp(V^n)}{(exp(V_j^e) + exp(V_k^e))^2} \right] = \sum_{k \in \mathcal{E}} \frac{-g_k}{(exp(V_j^e) + exp(V_k^e))^2} < 0.$$

Hence, for small enough V^n equation (101) is positive for all $exp(V_j^e)$ so that PR_j^m is increasing in $exp(V_j^e)$.

The intuition of the result is that "better firms hire from better firms," where nonemployment is viewed as an exceptionally bad firm.

The second approach looks at worker outflows and ranks firms based on the separation rate. This approach follows a long tradition in the inter-industry wage differential literature of using a survey-based measure of the quit rate as a measure of desirability (e.g., Ulman (1965, Table III) and Krueger and Summers (1988, Table IX)). The model offers several ways of operationalizing this idea, which hinges on how to interpret the survey response of "quit." One possibility is to interpret this as all EE transitions, which gives rise to the following pair of definitions:

$$(103) \qquad QR_j(EE)^o = \frac{\sum_{k \in \mathcal{E} \backslash j} M_{kj}^o}{g_j W}; \; QR_j(EE)^m = \frac{\sum_{k \in \mathcal{E} \backslash j} M_{kj}}{(1 - \delta)(1 - \rho) g_j W}.$$

Alternatively, the quit rate could be interpreted as all separations, which gives rise to the following pair of definitions:

$$(104) \qquad QR_j(ALL)^o = \frac{\sum_{k \in \mathcal{E} \cup n \backslash j} M_{kj}^o}{g_j W}; \; QR_j(ALL)^m = \frac{\sum_{k \in \mathcal{E} \cup n \backslash j} M_{kj}}{(1 - \delta)(1 - \rho) g_j W}.$$

Result 3 $QR_j(ALL)^m$ and $QR_j(EE)^m$ are monotonically decreasing in V_j^e.

Proof EE quit rate: The probability of an EE quit is given by (for simplicity, this includes the probability of a worker at firm j quitting to firm j):

$$(105) \qquad QR_j(EE)^m = \lambda_1 \sum_k f_k \frac{exp(V_k^e)}{exp(V_k^e) + exp(V_j^e)}.$$

Taking the derivative with respect to $exp(V_j^e)$:

$$(106) \qquad \frac{\partial QR_j(EE)^m}{\partial exp(V_j^e)} = \lambda_1 \sum_{k \in \mathcal{E}} f_k \frac{-exp(V_k^e)}{(exp(V_k^e) + exp(V_j^e))^2} < 0.$$

Hence, $QR_j(EE)^m$ is decreasing in $exp(V_j^e)$.

EN quit rate: The probability of an EN quit is given by:

$$(107) \qquad QR_j(EN)^m = (1 - \lambda_1)\frac{exp(V^n)}{exp(V^n) + exp(V_j^e)}.$$

Taking the derivative with respect to $exp(V_j^e)$:

$$(108) \qquad \frac{\partial QR_j(EN)^m}{\partial exp(V_j^e)} = (1 - \lambda_1)\frac{-exp(V^n)}{(exp(V^n) + exp(V_j^e))^2} < 0.$$

Hence, $QR_j(EN)^m$ is decreasing in $exp(V_j^e)$.

All quit rate: This result follows form combining the previous two results.

The intuition of the result is that because workers at all firms face the same offer distribution and, in expectation, value all firms the same way, the probability of choosing to leave is decreasing in the quality of the firm.

K Appendix: Inverting the value function

Following Hotz and Miller (1993), take advantage of two properties of Type I extreme value errors,[19] and to keep notation compact, use $p_{jk} = \Pr(j \succ k) = \frac{exp(V_j^e)}{exp(V_j^e) + exp(V_k^e)}$. Rearranging equation (??):

$$
\begin{aligned}
v_j = V_j^e &- \beta\mathbb{E}\{\delta\{V^n + \gamma\} + \rho_j(1 - \delta_j)\int_k \{V_k^e + \gamma\}\tilde{f} \\
&+ (1 - \rho_j)(1 - \delta_j) \times [(1 - \lambda_1)\{p_{nj}(V^n + \gamma - \ln p_{nj}) + p_{jn}(V_j^e + \gamma - \ln p_{jn})\} \\
(109) \qquad &+ \lambda_1\int_k \{p_{kj}(V_k^e + \gamma - \ln p_{kj}) + p_{jk}(V_j^e + \gamma - \ln p_{jk})\}f]\}.
\end{aligned}
$$

Solving this equation requires two objects that are not required to solve for V_j^e: β and $\tilde{f}$. I set $\beta = 0.95$ (reflecting the annual frequency of the model), and I set $\tilde{f} = f$. If there were no variation in δ_j and ρ_j across j, then v_j is just a monotone transformation of V_j^e since the first two terms do not vary in V_j^e and the second two terms are monotone increasing in V_j^e. With variation in δ_j and ρ_j, then this equivalence breaks. Nonetheless, the correlation between V_j^e and v_j is 0.937. Similarly, the correlation between v_j and Ψ_j is 0.573 (because it is computationally expensive to compute v_j, I did not compute this quantity in each bootstrap repetition and so this should be compared to the "raw" correlation between V_j^e and Ψ_j, which is 0.530.)

L Appendix: Additional tables and figures

---

[19]First, $\gamma = \mathbb{E}[\iota] \approx 0.577...$ is Euler's constant. Second, the conditional expectation of ι is related to the choice probability in the following way, where the two choices are a and b: $\mathbb{E}[\iota_a|a] = \gamma - \ln(\Pr(a \succ b))$.

Table A1: Constructing Sample of Dominant Jobs

	Number (1)	Unique People (2)	Unique Employers (3)
Person-employer-year pre-earnings test	$650,288,000$	$108,002,000$	$6,688,000$
Person-employer-year post-earnings test	$613,341,000$	$105,921,000$	$6,511,000$
Person-years	$504,945,000$	$105,921,000$	$6,155,000$

Notes: All counts are rounded to the nearest thousand. Row 2 divided by row 3 is 1.215. The first row shows the total number of person-year-employer observations that are continuous quarter or full-quarter among workers in the relevant age range. The second row shows the number of person-year-employer observations where the person's dominant job in the particular year passes an earnings test. The third row goes down to the unique employer that provides the worker's "dominant" job, or the employer from which the worker made the most in the calendar year.

Table A2: Distribution of jobs per person per year

	Number of person-years
1	$413,553,000$
2	$77,735,000$
3	$11,611,000$
4+	$2,047,000$

Notes: All counts are rounded to the nearest thousand. This table deconstructs the gap between row 2 and row 3 in Table A1. The column sum is row 3 in Table A1. This shows among workers in the sample of workers with dominant jobs the distribution of the number of continuous and full-quarter jobs in a year.

Table A3: Type of earnings in the annual dominant job dataset

Type of earnings	Number of person-years
Full quarter	$458,017,000$
Continuous quarter	$46,928,000$
Continuous quarter share	0.093

Notes: All counts are rounded to the nearest thousand. The column sum is the number of person-years in row 3 in Table A1. A worker is employed full-quarter in quarter t if she has earnings from her employer in quarter t and quarters $t-1$ and $t+1$. A worker is employed in a continuous quarter way in quarter t if she has earnings from her employer in quarter t and quarter $t-1$ or quarter $t+1$.

Table A4: Number of years per person

	Number of people	Share
1	14,041,000	0.133
2	11,422,000	0.108
3	9,873,000	0.093
4	9,111,000	0.086
5	8,963,000	0.085
6	10,396,000	0.098
7	42,115,000	0.398

Notes: All counts are rounded to the nearest thousand. The column sum is the number of unique people in row 3 in Table A1.

Table A5: Dominant employers per person

Number of dominant employers	Number of people	Share of people
1	52,938,000	0.500
2	27,228,000	0.257
3	14,945,000	0.141
4	7,157,000	0.068
5	2,764,000	0.026
6	771,000	0.007
7	118,000	0.001

Notes: All counts are rounded to the nearest thousand. The column sum is the number of unique people in row 3 in Table A1.

Table A6: Number of years per match

Years per match	Matches (person-employers)	Share of matches	Share of person-years
1	93,327,000	0.466	0.185
2	39,176,000	0.196	0.155
3	19,842,000	0.099	0.118
4	12,295,000	0.061	0.097
5	8,573,000	0.043	0.085
6	6,745,000	0.034	0.080
7	20,175,000	0.101	0.280

Notes: All counts are rounded to the nearest thousand. The column sum in the first column is the number of matches and is approximately 200,000,000, which is between the number of unique people and the number of person-years. The next column shows the distribution by share of matches. The last column shows the distribution of person-years.

Table A7: Composition of separations in the quarterly dataset

Type of transition	Definition	Number
employer-to-nonemployment	Standard	131,621,000
employer-to-employer	Standard	76,152,000
employer-to-employer	New	2,680,000
employer-to-employer transition share		0.375
New definition share		0.035
Total separations		210,453,000

Notes: All counts are rounded to the nearest thousand. The dataset is the quarterly dataset, so it includes some workers not in the annual dataset. The standard definition uses overlapping quarters to measure employer-to-employer transitions. The new definition uses stability of earnings to measure employer-to-employer transitions.

Table A8: Summary statistics and the variance of earnings with the selection-correction

	S. Connected by EE
Sample size	
People-years	$409,550,000$
People	$90,895,000$
Employers	$476,000$
Summary statistics	
Mean log earnings	10.48
Variance of log earnings	0.67
Share of variance of earnings explained by each parameter set	
Employers	0.21
People	0.57
Xb	0.11
Selection-correction	0.00
Variance components	
Variance of emp. effect	0.14
Variance of person effect	0.50
Variance of Xb	0.07
Variance of selection-correction	0.00
2cov(person, emp.)	0.10
2cov(Xb, person + emp.)	0.08
2cov(selection-correction,person+emp.)	0.00
2cov(selection-correction, Xb)	0.00
Corr(person, emp.)	0.19
Overall fit of AKM decomposition	
Adj. R^2	0.86
Match effects model	
Adj. R^2	0.92

Notes: Sample counts are rounded to the nearest thousand. The data is at an annual frequency. There is one observation per person per year. The observation is the job from which a person made the most money, but only if she made at least $3,250 (in $2011, using the CPI-U). Earnings are annualized. The table includes person-years in which on December 31 the person was aged 18-61 (inclusive). EE is employer-to-employer. The sample is the same as column (3) of Table **??**.

Table A9: Hours and compensating differentials

|	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Variation in hours across sectors							
Agriculture	N/A	N/A	44.7	44.20	45.60	46.40	46.70
Mining	49.04	49.20	49.20	41.00	41.40	41.50	41.60
Construction	40.70	40.90	40.90	41.10	41.40	41.50	41.70
Manufacturing	42.36	42.30	42.30	42.20	42.60	42.60	42.80
Wholesale	38.22	38.10	42.60	38.60	39.00	39.70	40.40
Retail	38.22	38.10	36.90	37.80	38.20	39.00	39.60
Transport and Warehousing	42.20	42.30	42.20	42.30	42.80	42.80	42.90
Utilities	42.20	42.30	42.30	40.70	41.00	41.40	41.70
Information	39.88	39.60	39.60	40.20	40.50	40.80	41.00
FinInsurance	40.24	40.00	40.40	40.10	40.50	40.60	40.70
Real Estate	40.24	40.00	39.10	40.80	41.40	41.50	41.80
ProfSciTech Services	40.16	40.00	41.20	40.20	40.70	40.80	41.20
Management	40.16	40.00	42.70	37.10	37.60	37.80	38.40
Admin and Waste	40.16	40.00	38.20	37.50	37.80	38.00	38.50
Education	37.38	37.30	36.90	37.50	37.70	38.00	38.40
Health and Social	37.38	37.30	37.50	34.60	34.40	36.20	37.30
Arts and Rec	34.36	34.20	34.80	34.00	33.80	35.60	36.60
Accom and Food Serv	34.36	34.20	34.00	36.60	37.20	37.70	38.40
Other Services	36.88	36.70	36.70	40.00	40.50	40.70	41.00
Public Administration	40.84	40.70	40.70	40.20	40.30	40.40	40.60
N	625741	625739	681264	654715	467121	459650	445962
Panel B. Relationship between hours and compensating differentials							
R^2	0.323	0.325	0.307	0.149	0.158	0.160	0.169

Notes: Panel A report average hours worked last week across sector pooling data from 2003-2007. Columns (1) through (3) uses the monthly CPS. Column (1) reproduces the weighted average of the published BLS Annual Average tables (Table 21). The BLS tables are more aggregated than the sector level, and so the numbers are copied where two sectors are joined (for example, "education" and "health and social" are aggregated). Column (2) replicates column (1) using the micro-data. Column (3) disaggregates the sectors that were combined in columns (1) and (2), and adds in agricultural workers. Columns (4) through (7) use the March CPS. Column (4) imposes the same sample restrictions as in columns (2) and (3). Column (5) adds sample restrictions following Autor et al. (2008) in terms of dropping imputed observations and observations with extreme earnings. Column (6) restricts to men and women aged 18-61. Finally, column (7) imposes an earnings floor of $3250 a year (in $2011). Panel B reports the relationship between estimated sectoral compensating differentials reported in Figure ?? and measures of the variation in hours worked across sectors reported in each column in Panel A. The R^2 comes from a regression of the sectoral compensating differentials on the the hours reported in the relevant column, where the observations are weighted by the sectoral sum of the relevant person weights.

Table A10: Monte Carlo Results

	One in 10 sampling				One in 20 sampling				One in 200 sampling			
Panel A. True minus estimated												
	Mean	50th	10th	90th	Mean	50th	10th	90th	Mean	50th	10th	90th
Raw R^2 of V^e and Ψ	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
Split sample adjusted R^2 of V^e and Ψ	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	-0.01	0.00
Bootstrap adjusted R^2 of V^e and Ψ	NA	NA	NA	NA	NA	NA	NA	NA	0.00	0.00	0.00	0.01
Raw large firm R^2 of V^e and Ψ	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.01
Level of true R^2 of V^e and Ψ	0.26	0.26	0.26	0.26	0.32	0.32	0.32	0.32	0.32	0.32	0.31	0.32
Level of true R^2 of V^e and Ψ at large firms	0.33	0.33	0.33	0.33	0.43	0.43	0.43	0.43	0.55	0.55	0.55	0.55
Firm share of variance of earnings (raw)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Firm share of variance of earnings (large firm)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Level of true firm share	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
Level of true firm share (large firm)	0.22	0.22	0.22	0.22	0.24	0.24	0.24	0.24	0.22	0.22	0.22	0.22
Panel B. Correlations between true and estimated values												
Ψ	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
V^e	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99
$\tilde{V}^{EE}$	0.81	0.81	0.81	0.82	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.84

Notes: This table reports Monte Carlo simulations of the estimation procedure. The table reports statistics across 100 simulation runs. The left-hand side reports simulations where I have drawn one-in-10 random sample; the middle panel reports a one-in-20 random sample and the right-hand side reports one-in-200 random sample. Because the bootstrap is very computationally expensive, I only report the bootstrap results for the one-in-200 sample. Panel A reports statistics on the gap between the true minus the estimated value, while Panel B reports statistics on the correlation between the true and estimated values. The interpretation of a positive number in Panel A is that the estimation procedure underestimates the true relationship. The definitions and procedures are the same as in Table **??**.
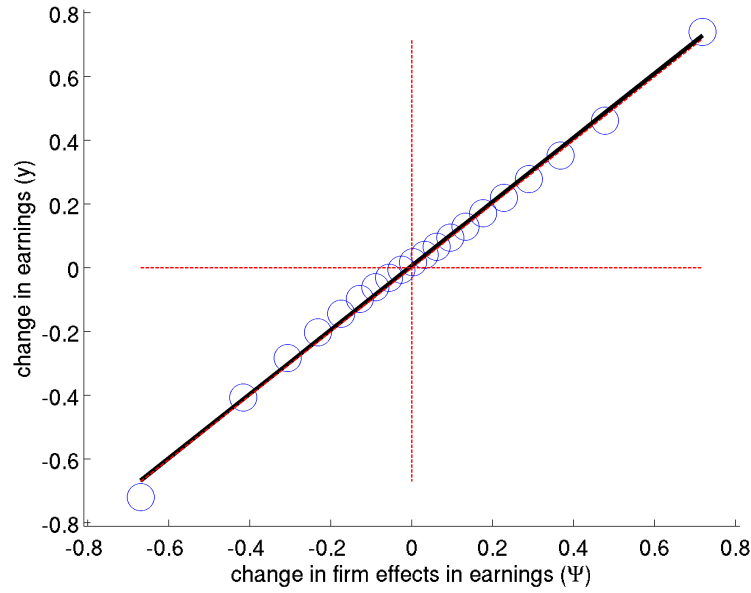
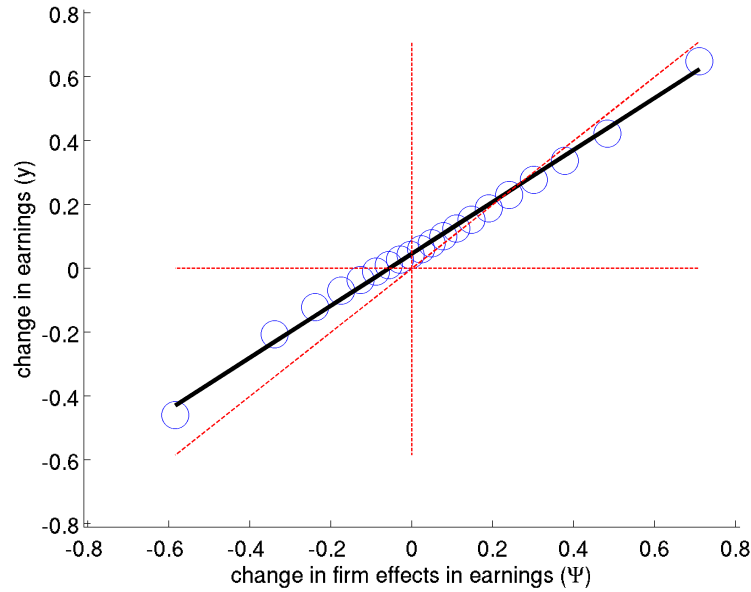Figure A1: States used in analysis
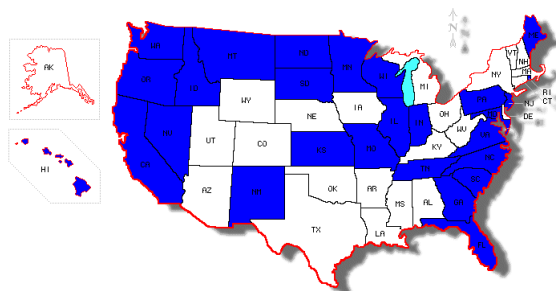
Notes: The states in blue are used in the analysis.

Figure A2: Change in firm pay related to magnitude of earnings change

(a) All

(b) EE

Notes: These figures show how the magnitude of earnings changes relate to the change in firm-level pay for workers who switch annual dominant jobs. The earnings are the residualized annualized earnings in the last year at the previous job and in the first year at the new job. The top panel looks at all transitions and the bottom panel looks at employer-to-employer (EE) transitions. I sort the job changers into 20 bins on the basis of the change in the firm effects. The circles plot the bin means. The solid line plots the best-fitting line estimated based on the micro-data. The dashed red line plots the 45 degree line. The coefficient in the upper panel is 1.005 (standard error: 0.0003), and in the bottom panel is 0.813 (standard error: 0.0003).

Figure A3: Change in firm effect does not predict magnitude of earnings change in a matching model
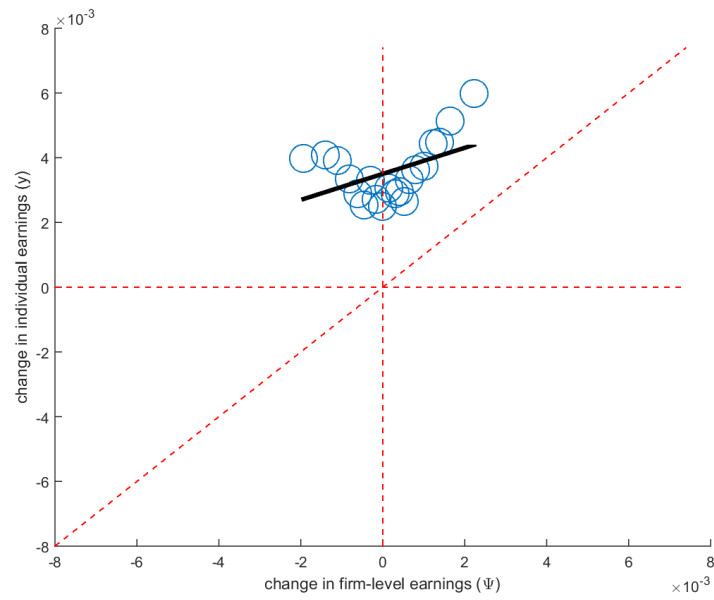
Notes: This figure is based on simulating the example production function in Eeckhout and Kircher (2011) and is constructed in a manner analogous to Figure A2.

Figure A4: Event studies of earnings changes
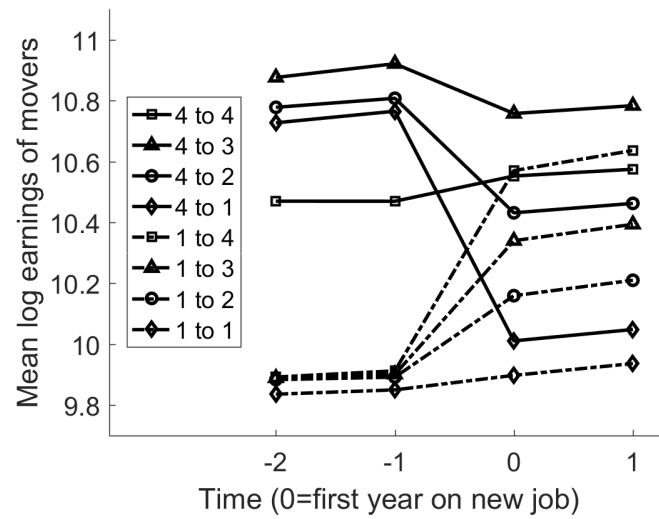
Notes: This figure shows the mean wages of workers who change jobs and held the preceeding job for two or more years, and the new job for two or more years. "Job" refers to the dominant annual job. Each job is classified into quartiles based on the estimated firm effects in Table **??** column (3).

References

Abowd, John M., Paul Lengermann, and Kevin L. McKinney, "The Measurement of Human Capital in the U.S. Economy," Working Paper 2003.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 90 (2008), 300–323.

Bagger, Jesper and Rasmus Lentz, "An Equilibrium Model of Wage Dispersion and Sorting," *Review of Economic Studies*, (2017).

Benedetto, Gary, John Haltiwanger, Julia Lane, and Kevin Mckinney, "Using Worker Flows to Measure Firm Dynamics," *Journal of Business and Economic Statistics*, 25 (2007), 299–313.

Bjelland, Melissa, Bruce Fallick, John Haltiwanger, and Erika McEntarfer, "Employer-to-Employer Flows in the United States: Estimates Using Linked Employer-Employee Data," *Journal of Business and Economic Statistics*, 29 (2011), 493–505.

Burgess, Simon, Julia Lane, and David Stevens, "Job Flows, Worker Flows, and Churning," *Journal of Labor Economics*, 18 (2000), 473–502.

Card, David, Joerg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128 (2013), 967–1015.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Health Care Exceptionalism? Performance and Allocation in the U.S. Health Care Sector," *American Economic Review*, 106 (2016), 2110–2144.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104 (2014), 2593–2632.

de Melo, Rafael Lopes, "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence," *Journal of Political Economy*, (2016).

Eeckhout, Jan and Philipp Kircher, "Identifying Sorting—In Theory," *Review of Economic Studies*, 78 (2011), 872–906.

Hagedorn, Marcus, Tzuo Hann Law, and Iourri Manovskii, "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85 (2017), 29–65.

Hotz, V. Joseph and Robert A. Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60 (1993), 497–529.

Hyatt, Henry R. and Erika McEntarfer, "Job-to-Job Flows and the Business Cycle," Working Paper 2012.

— , — , Kevin McKinney, Stephen Tibbets, and Doug Walton, "Job-to-Job Flows (JTJ) Flows: New Labor Market Statistics From Linked Employer-Employee Data ," Working Paper CES 14-34 2014.

Krueger, Alan B. and Lawrence H. Summers, "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, 56 (1988), 259–293.

Minc, Henryk, *Nonnegative matrics*, Wiley, 1988.

Morris, Carl N., "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78 (1983), 47–55.

Taber, Christopher and Rune Vejlin, "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," Working Paper 2016.

Ulman, Lloyd, "Labor Mobility and the Industrial Wage Structure in the Postwar United States," *Quarterly Journal of Economics*, 79 (1965), 73–97.